Delft University of Technology

Telepresence for surgical assistance and training using eXtended reality during and after pandemic periods

Wisotzky, Eric L.; Rosenthal, Jean Claude; Meij, Senna; van den Dobbelsteen, J.J.; Arens, Philipp; Hilsmann, Anna; Eisert, Peter; Uecker, Florian Cornelius; Schneider, Armin

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Telepresence for surgical assistance and training using eXtended reality during and after pandemic periods

Eric L Wisotzky[1,2,3,*] (iD), **Jean-Claude Rosenthal**[1,2,*], **Senna Meij**[4], **John van den Dobblesteen**[4], **Philipp Arens**[5] (iD), **Anna Hilsmann**[1], **Peter Eisert**[1,3] (iD), **Florian Cornelius Uecker**[5], and **Armin Schneider**[6,7] (iD)

Abstract
Existing challenges in surgical education (See one, do one, teach one) as well as the COVID-19 pandemic make it necessary to develop new ways for surgical training. Therefore, this work describes the implementation of a scalable remote solution called "TeleSTAR" using immersive, interactive and augmented reality elements which enhances surgical training in the operating room. The system uses a full digital surgical microscope in the context of Ear–Nose–Throat surgery. The microscope is equipped with a modular software augmented reality interface consisting an interactive annotation mode to mark anatomical landmarks using a touch device, an experimental intraoperative image-based stereo-spectral algorithm unit to measure anatomical details and highlight tissue characteristics. The new educational tool was evaluated and tested during the broadcast of three live XR-based three-dimensional cochlear implant surgeries. The system was able to scale to five different remote locations in parallel with low latency and offering a separate two-dimensional YouTube stream with a higher latency. In total more than 150 persons were trained including healthcare professionals, biomedical engineers and medical students.

Introduction

In 2020, the COVID-19 pandemic dramatically changed medical care in many ways. It has been a driver for digitization in medicine,[1] however, there have been severe cutbacks for surgical training due to contact restrictions and emergency services in hospitals.[2] Nonetheless, digital treatment and consultation methods are being implemented quickly.[3] More and more patients and physicians are using video consultations or other digital applications in their daily routine. In 2017, only 2% of physicianss used video consultation in Germany, whereas in 2020, there was a strong increase to 50% followed by an additional increase of 10% in early 2021.[4,5] This rapid development is driven by two reasons: First, video consultation is a fast and effective way to protect physicians and patients during the pandemic while maintaining medical consultation. Second, video consultation adds a visual layer

[1]Fraunhofer Heinrich-Hertz-Institute HHI, Berlin, Germany
[2]Klinik und Poliklinik für Hals-Nasen-Ohrenheilkunde, Kopf- und Halschirurgie "Otto Körner", Rostock University Medical Center, Rostock, Germany
[3]Department of Informatics, Humboldt University, Berlin, Germany
[4]Delft University of Technology, Faculty of Mechanical Engineering, BioMechanical Engineering, Delft, The Netherlands
[5]Department of Otorhinolaryngology, Charité – Universitätsmedizin Berlin, Berlin, Germany
[6]Munich Surgical Imaging, Munich, Germany
[7]Department of Engineering Sciences, Jade Hochschule, Wilhelmshaven, Germany

*These authors contributed equally to this work.

Corresponding author:
Eric L Wisotzky, Fraunhofer Heinrich-Hertz-Institute HHI, Vision and Imaging Technologies, Einsteinufer 37, Berlin 10587, Germany.
Email: eric.wisotzky@hhi.fraunhofer.de

compared to phone calls making important medical correlations easier to explain.

This positive development contrasts with existing limitations of surgical training following the classical teaching paradigm "See one, do one, teach one."[6–8] These limitations become even more critical in pandemic scenarios when physicians cannot be trained to the usual extent due to contact restrictions in the operating room (OR) or canceled routine interventions. To address these limitations, simulation-based training has been proposed as a method of medical education.[9–11] However, surgical training requires the acquisition of extensive knowledge of all surgical steps and a minimum number of operations under supervision. Even during the COVID-19 pandemic, students reported that ''real cases'' are preferred for training instead of ready-made cases.[12] Additionally, there is a lack of training measures, pedagogical, and technical education to reduce the digital illiteracy of physicians and prepare them for the 'digital OR' in the upcoming decade.[13–16]

The usage of eXtended reality (XR) such as augmented reality (AR), virtual reality (VR), and mixed reality (MR) for training has been surveyed over the last years[17–19] and shows potential in many ways.[20–23] It can improve the surgical knowledge[24,25] and help surgeons to extend the limited field-of-view during endoscopic surgeries.[26–28] However, the use of live streaming technologies for surgical training is quite new with only a few concepts arise in the last eight years.[29] These concepts only focused on streaming the surgical situation using an external camera allowing the trainee an insight into the OR, but already showed a positive impact on the learning outcome[29] without enlarging patient risks.[30]

In microscopic surgery, another challenge for continuous medical education (CME) presents itself: it requires the same view identically to the surgeon's view. Currently, CME is structured as one-sided training in groups of up to 10 persons.[31] However, due to resource constraints (time, limited space) and hygienic restrictions in the OR the direct "surgical view" through the microscope cannot be delivered continuously to all trainees. This can lead to an inconsistent and slightly different training level.[32] One of the missing features is the fundamental three-dimensional (3D) visualization of microscopic procedures; trainees typically follow the operation via a camera attached to the microscope and connected to a 2D/3D display resulting in a different field-of-view.[29] Therefore, the surgeon's view is conveyed to others to a limited extent. In this case, depth impression and relationships of individual anatomical structures are lost and the learning effect is significantly reduced. Additionally, it is a great challenge for surgeons in training to differentiate similar-looking tissue structures correctly, for example, risk structures like the facial nerve or pathological structures like proliferating cholesteatoma tissue.[33,34]

The overall objective of this work is the implementation of a scalable, real-time capable audio and video processing chain using an AR-based stereo-spectral algorithm unit for on-site and remote surgical training and education. The remainder of this article is organized as follows: The next section describes the overall AR system design including the stereo-spectral algorithm unit generating important anatomical information. In the section "Results," three conducted ENT courses are evaluated with respect to lessons learned and user feedback. The last section gives an outlook for the future of surgical training and how TeleSTARs features can be used for intraoperative decision-making.

Materials and methods

Scalable remote system design

The overall system is designed for a low latency bi-lateral audio and XR-based 3D video pipeline supporting full surgical transparency in remote scenarios:

1. **AR-based 3D video pipeline**, sharing the exact same surgical view to multiple participants in stereoscopic 3D following the digital twin paradigm.
2. **Bi-directional communication**: A permanent audio connection and interactive visual communication interface.
3. **Real-time network configuration**, guaranteeing a low latency, secured, and synchronized transmission of the surgery.

Scalability to other remote locations is achieved by connecting additional lecture rooms using the outlined hardware and network infrastructure, see Figure 1. However, the described system concept relies on a well-working low latency internet infrastructure offering a high bandwidth and low round-trip times (<30 ms). In addition to directly connect other remote locations in 3D, it is possible to provide a 2D-YouTube-Livestream using our proposed setup.

AR-based 3D surgical video pipeline. The overall system design is depicted in Figure 1. It consists of four core components using dedicated hardware and software interfaces.

(1) A full digital surgical microscope[i] (DSM) having an extendible software XR interface[21] (Figure 1 items A to E).

(2) An external 2D overview camera presenting the OR setup and periphereal activities (Figure 1F).

(3) A multichannel audio/video mixer,[ii] which fuses all streams (Figure 1G).

(4) A real-time H.264 video encoder in the OR (Figure 1H) and H.264 video decoders[iii] at remote locations (Figure 1I) receiving the AR-based video stream with audio.

All remote locations are connected to the OR over a local area network (LAN) or a wide area network (WAN). Due to the systems scalability, trainees can watch the video on 3D
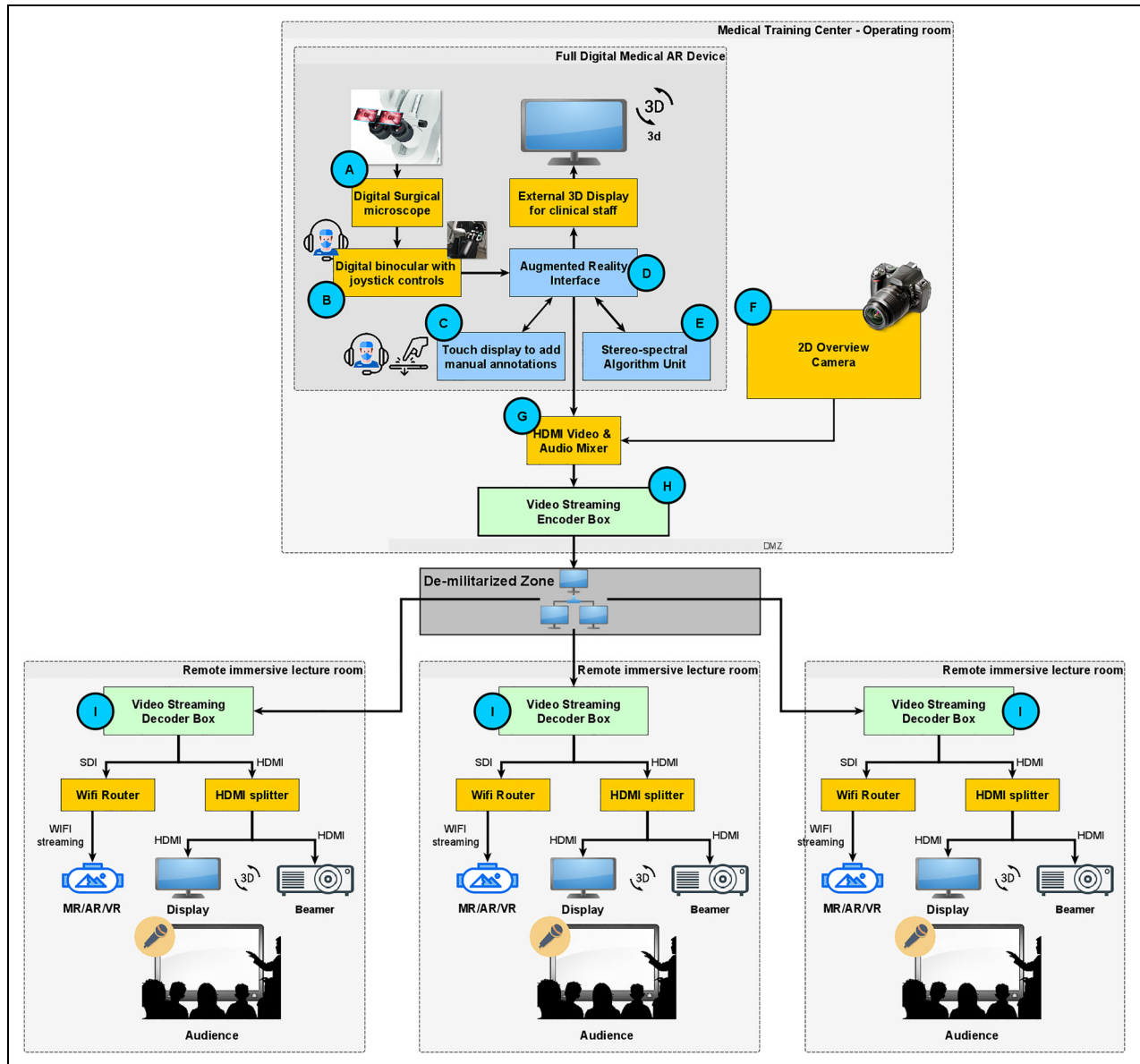
Figure 1. Scalable AR-based 3D system design for remote surgical training and education. AR: Augmented Reality; 3D: three-dimensional.

displays/projectors or XR-based head-mounted-displays (HMDs), which can be connected to HDMI/SDI interfaces of the video decoders. The number of remote locations, that is, video decoders is not restricted and has no effect on video quality nor the representation of the AR features. Thus, an unlimited scalability is guaranteed.

Bi-directional audio communication. Besides the AR-based 3D video pipeline, the system has a bi-directional audio pipeline for real-time communication between trainers and trainees. To fulfill these requirements the system has an audio commentary concept avoiding disturbing echo and latency effects.

In the OR, we set up a forward audio channel consisting of two audio inputs: the audio track of the surgeon is directly connected to the video pipeline of the microscope, synchronized and embedded in the underlying HDMI stream (Figure 1B). The moderating surgeon uses a Bluetooth headset and a wireless microphone (Figure 1C). Both audio streams are fed into the audio/video mixer pipeline as depicted in Figure 1G. The mixer allows switching between both audio channels and muting if needed. The audio back channel for remote questions is implemented using a conferencing tool to which the headset of the moderating surgeon is connected. This surgeon moderates and answers or forwards questions
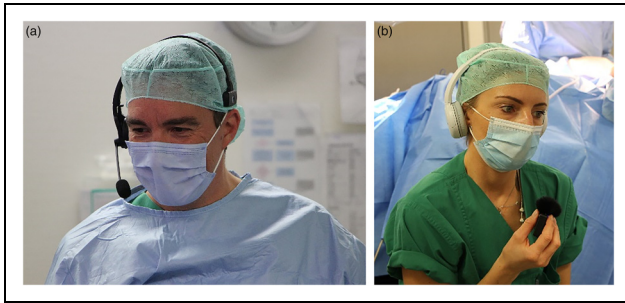
Figure 2. Audio system design for bi-lateral communication in a remote surgical training environment. (a) Operating surgeon wearing a Bluetooth headset explaining the intervention. (b) Moderating surgeon using a Bluetooth headset and a wireless microphone receiving questions from the remote audience.

Figure 3. Network configuration: firewall and de-militarized zone.

Figure 4. Annotation mode of live image. The blue border around the image indicates that the augmentation mode is activated.

Figure 5. Illumination spectra of the four LED.

to the operating surgeon when possible. The operating surgeon only uses the microphone to minimize disturbing background noise and receives feedback from the moderating surgeon (Figure 2(a) and (b)).

Network configuration. Low latency and secure data transmission are guaranteed by a demilitarized zone (DMZ). The DMZ firewall allows only connections from known external IP addresses (Figure 3). The DMZ is implemented using a virtual server including a reverse proxy, which is administered and monitored through secure VPN connections. The reverse proxy guarantees a secure connection between the internal streaming server inside LAN (Figure 1H) and the external remote WAN clients (Figure 1I), handling the 3D video decoding process.

Intraoperative tools

TeleSTARs intraoperative toolchain has three parts: (1) an annotation tool, (2) a 3D reconstruction pipeline allowing image-based measurements,[35] and (3) a multispectral analysis module allowing tissue differentiation.[36] All XR-features are beneficial for intraoperative assistance and surgical training.

Annotation tool. The DSM's touch-screen user-interface (UI) allows direct and intuitive annotations of the surgical scene. The annotations are fused into the image shown in the binocular. Due to the complexity of rendering virtual objects into correct depth layers, the annotation is performed in 2D only (left view). During annotation, the microscopic head is static with fixed brakes. Any release of the brakes or change of magnification or focus deletes the annotated information to ensure the correctness of earlier annotated structures. The annotation tool (Figure 4) has six modes: (1) cross marker, (2) circle, (3) boxing, (4) directional
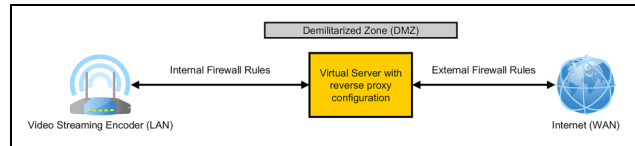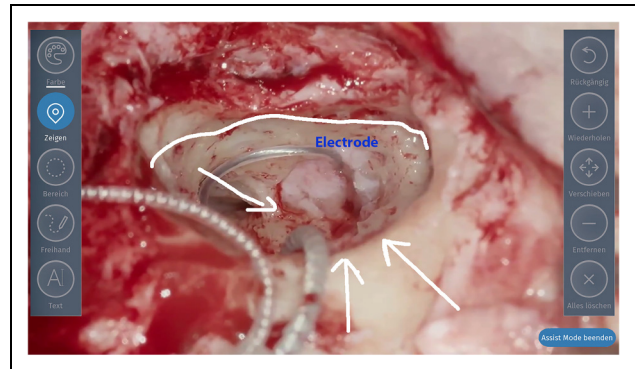
arrow, (5) free-hand, and (6) text. All annotations can be colored, edited, or deleted.

Multispectral analysis. The DSM has an RGB sensor and an LED light source. The multispectral analysis unit uses the integrated LEDs which are synchronized to the RGB sensor frequency allowing to capture a sequence of $N = 4$ consecutive different illuminated images. The compound light source is a cluster of "warm white", "red", "green", and "cold white" LEDs, see Figure 5. The spectral response of the three channels ($M = 3$) is presented by Liu et al.,[37] Leonhardt and Brendel,[38] and Clark et al.[39]

Hence, 12 spectral images can be recorded by using the LED sequence combined with the RGB channels. Each spectral image captures information about specific spectral characteristics, which recur differently in different spectra, see Figure 6.

To split up these characteristic peaks into P independent bands, a non-negative linear least-squares problem is solved

$$\arg\min_x \|\mathbf{A}x - y\|_2 \qquad (1)$$

subject to $x = x' - x''$, where $x' \leq 0$ and $x'' \leq 0$. The matrix $\mathbf{A}$ holds the 12 spectra $(N \times M)$ and y is a vector. For each peak an optimization step is performed resulting in a correction matrix $\mathbf{C}$ with size of $(N \cdot M) \times P$.

Thus, a sequence can be combined into a spectral data cube, where each spatial pixel is represented as a vector with a size of $(N \times M)$. This cube is corrected according to Wisotzky et al.[40] For each pixel, the reflectance at $P = 10$ characteristic peaks in the visual spectrum is reconstructed.

3D reconstruction. It is a crucial process for the next generation of intraoperative applications for microscopy and other surgical disciplines (e.g. visceral surgery). In the context of image-guided surgery and remote surgical education, it creates a true-to-scale 3D surface representation of the patient's anatomy to get an improved understanding while also allowing image-based measurement of anatomical landmarks. Our 3D reconstruction pipeline has been successfully evaluated and compared to others.[41]

Before 3D reconstruction can be applied, a calibration process of the stereoscopic system needs to be performed which estimates the DMS's optical lens parameters.[42] The automated principle for multiple zoom levels using a zoom-independent calibration chart (Figure 7) is depicted in Figure 8. Figure 9 shows the overall calibration process which has six main steps.

Surgical course

Surgical training today involves trainees looking directly at the situs or watching a 2D monitor. During non-critical situations in microscopic surgeries, they may look into the binocular to perceive the surgical scene in 3D. However, this is time consuming and can delay the surgery significantly. Hence, in critical situations teaching is continued only in a limited manner, although such situations are important to build up valuable surgical knowledge.

Courses for CME typically have 10 to 15 participants. Due to limited time and space resources, it is not possible that all trainees take a look into the binoculars, even the line-of-sight to the monitor could be blocked. These limitations have a negative effect on the understanding of the intervention resulting in a prolong learning curve and a reduced teaching success.

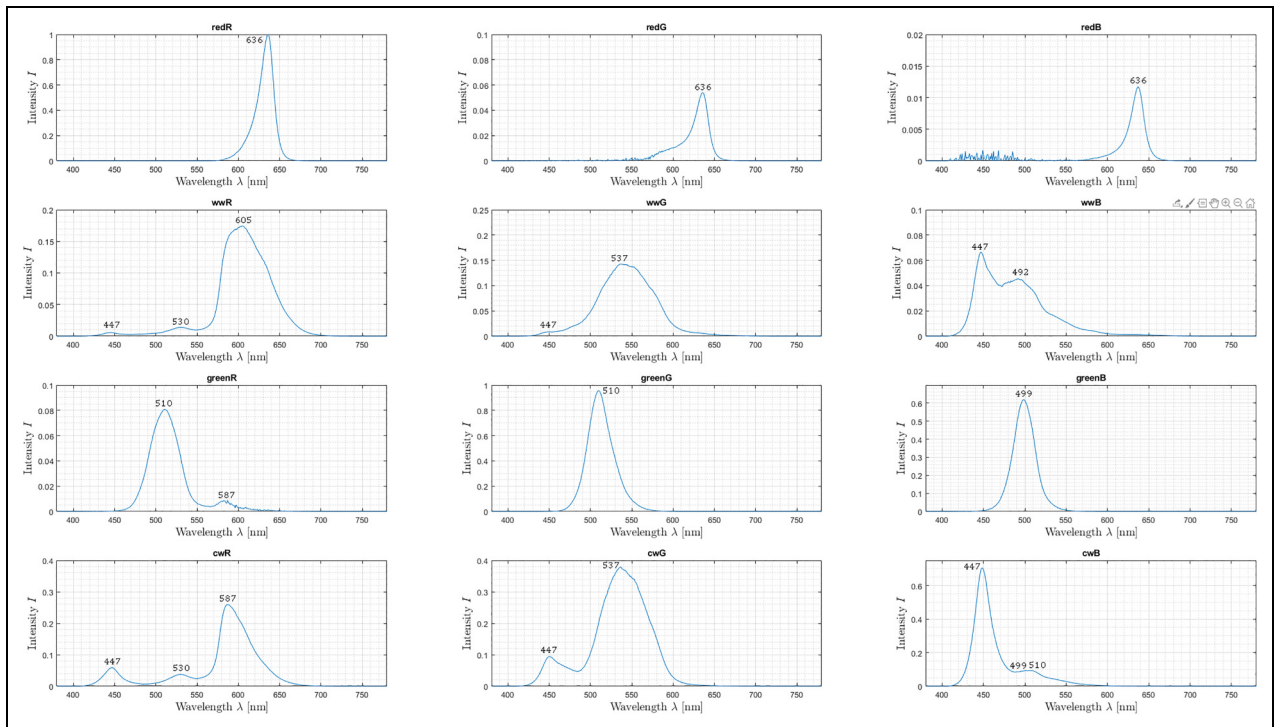
Figure 6. A spectral imaging sequence showing captured 12 spectral images. Relevant peaks in each spectrum are labeled with the corresponding wavelength λ position. In total, 10 wavelengths were selected.

Additional problems have arisen due to the COVID-19 pandemic in 2020. Many courses were canceled, leading to an obvious and significant delay in the training of clinical staff.[43] Therefore, we designed a hybrid course using XR under the highest hygienic standards and with approval of the pandemic staff of Charité – Universitätsmedizin Berlin and accompanied by the Medical Association of Berlin.

The lectures were performed in a conventional way followed by a question-and-answer session by different lecturers. All contents were streamed in 3D video and enriched with intraoperative information using the described tools. The XR-streams were sent to different remote lecture halls as well as 2D-stream to video platforms. During surgery, remote participants were able to interact with the moderating surgeon, for example, Q&A or sending images, 2D video platform participants could only send questions via chat.

Cochlear implant (CI). CI surgery was chosen for this study due to four reasons. (1) Complexity: as surgery at the lateral base of the skull it is difficult to learn and only a small number of experts exist. Moreover, its worldwide relevance is increasing.[44] (2) Highly standardized procedure: predictable intervention with duration of 75 min in average and very clear anatomical conditions. Depending on the conditions, timestamps rarely vary by more than ±5 min. Figure 10 shows five different timestamps (TSs) of the CI surgery storybook including TeleSTAR's XR features. (3) Landmarks: various tissue structures are exposed such as semicircular canals, bone, nerve, muscle, membranes, or silicone of the implant. (4) Mastoidectomy (Figure 10 Timestamp 2) as the mandatory part of the procedure can be practiced very well in XR environments or in a temporal bone workshops.[45,46] Beginners will benefit by perceiving the complexity of a CI on several levels. Experts can deepen their knowledge and discuss important steps directly with the teaching surgeon.

Results

Technical results

The individual results of the three intraoperative tools are presented, followed by the system performance.

The annotation tool allows an annotation of important tissue structures in different colors to highlight these for remote participants. In addition, the annotations can be augmented by text. The annotation tool was validated independently in an in-vivo study with participants of different training levels showing an improvement in learning outcome as well as in communication between trainee and trainer.[47] All augmentations are visualized instantly with no latency.

The multispectral unit analyzes the tissue in the situs allowing it to differentiate between single types. Due to the single selection of the individual LEDs, the illumination intensity of the single-LED-images is reduced by a factor of ~ 7.4. This results in a higher noise intensity. The intensity of the 12 spectral images differs due to illumination and sensitivity changes for the four LEDs and the RGB channels as shown in the signal-to-noise ratio (SNR) presented in Table 1. The SNR is calculated in a homogenous region R using
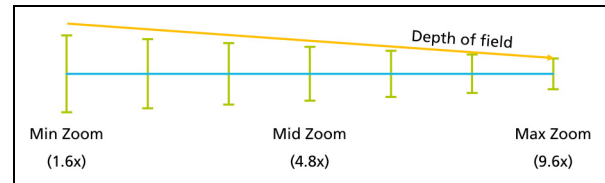
$$SNR = \frac{\bar{R}}{\sigma(R)} \quad (2)$$

Figure 8. Calibration strategy: motor-controlled capturing of different zoom levels.
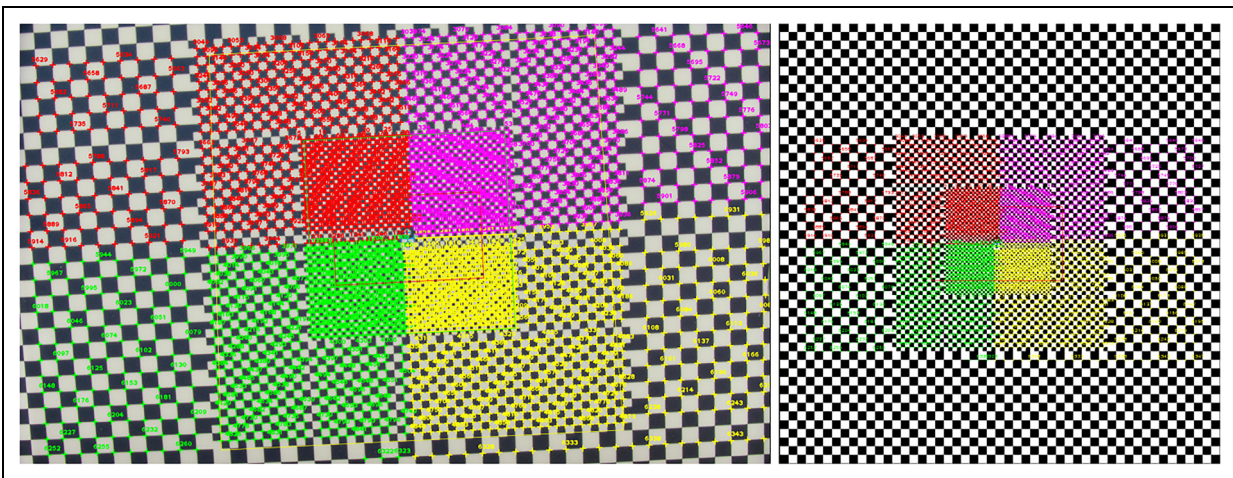

Figure 7. Intermediate calibration showing the color-encoded 2D to 3D correspondence mapping of detected features and 3D model features. Left: Detected features in four different quadrants. Right: Reference 3D model features of rendered model in canonical view. Each feature has a unique ID for a detailed 2D/3D evaluation. 2D: two-dimensional; 3D: three-dimensional.
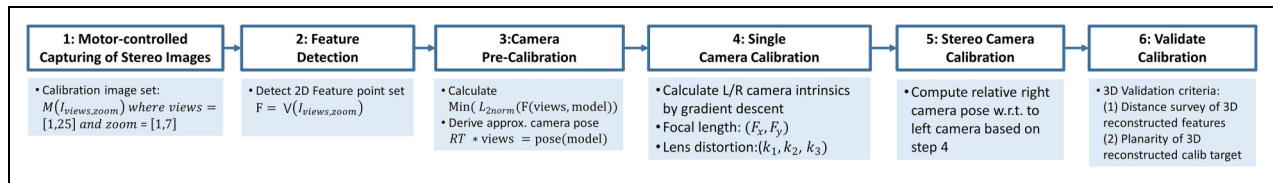
Figure 9. Calibration pipeline: motor-controlled capturing of different zoom levels with changing depth of field.
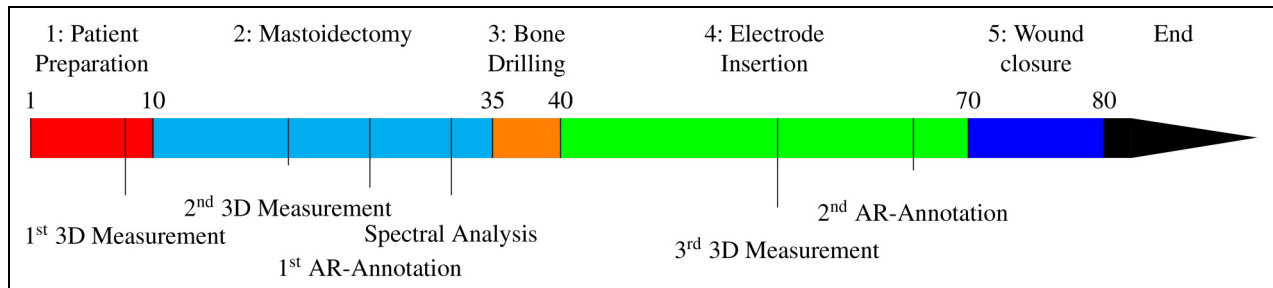

Figure 10. Timeline in minutes for a cochlear implant at surgery: Timestamps of intraoperative AR-features and annotation tools for remote surgical education. In the Appendix, all procedure steps are described in more details.

Table 1. Average signal-to-noise ratio (SNR) of the 12 spectral images.

Sensor channel	R	G	B
Red LED	6.75	1.77	1.01
Warm-white LED	5.58	3.81	1.64
Green LED	1.50	6.17	3.82
Cold-white LED	5.25	5.52	3.76

meaning $SNR = 1$ is the lowest possible value holding no information except noise.

Figure 11 gives an overview of all 12 spectral images, showing that sensor channel B holds no information when illuminating the scene with the red LED. This is expected behavior as the spectrum of red LED has a peak at 636 nm showing no overlap with the sensitivity spectrum of B channel. As introduced ten spectral bands are reconstructed. These reconstructed bands show good SNR (Table 2) and different spectral curves can be extracted for different tissue types.

The stereo system was calibrated for seven zoom levels. Calibration results are listed in Table 3. The accuracy for each zoom level is in the sub-millimeter range and is best for minimum zoom (<0.1 mm). Thus, the stereo image-based measurement unit fulfills the accuracy requirements for CI interventions and allows metric distance calculations between anatomical regions. In addition, Figure 12(a) and (b) shows a reconstructed true-to-scale point cloud of the scene while the CI is inserted.
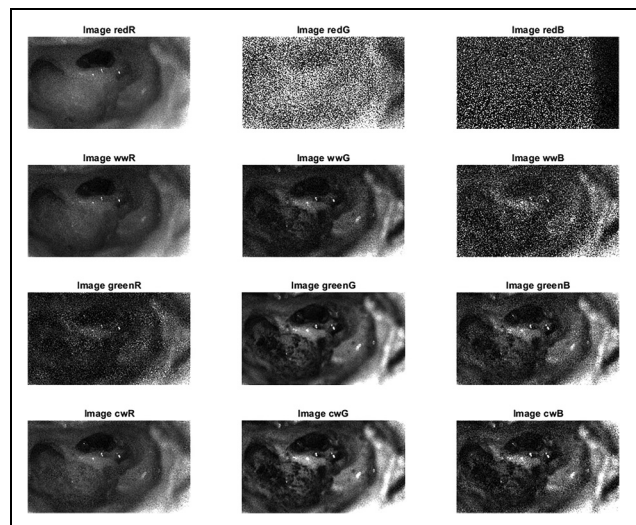
Figure 11. The 12 acquired spectral images of the third patient.

Runtime performance. The pipeline has three parts affecting the overall runtime performance: (1) the main digital video processing and transcoding pipeline, (2) the multispectral analysis module adds ~200 ms by the sequential toggle of the LEDs, and (3) the underlying network bandwidth. The transmission time of the surgery from the OR to the lecture rooms was about 600–700 ms depending on measured round-trip-times of ~20–30 ms and configured cache sizes on the remote end plus another 200–300 ms

Table 2. Average signal-to-noise ratio (SNR) of the 10 reconstructed wavelength images.

447 nm	492 nm	499 nm	510 nm	530 nm	537 nm	582 nm	587 nm	605 nm	636 nm
3.61	9.69	13.45	8.11	8.97	7.91	8.62	10.91	6.53	6.29

Table 3. Calibration evaluation: absolute mean errors of control points with known ground-truth for seven zoom levels.

Zoom level	Measurement accuracy [mm]	Planarity of reconstructed 3D points [mm]
1.58×	0.0947	0.00697
2.91×	0.0503	0.00551
4.25×	0.2123	0.00987
5.59×	0.4136	0.02625
6.92×	0.3518	0.02294
8.26×	0.3779	0.01887
9.60×	0.3114	0.01614

for the audio reverse channel into the OR. Hence, the total transmission time was slightly below one second in average allowing a seamless and interactive communication between the teaching surgeon and remote trainees as the tempo-spatial consistency was still good enough to follow the intervention. The importance of two common latency limits for surgical hand-eye coordination (50–80 ms) and for conferencing tools (200 ms) could be neglected in our case as the remote trainees watched a synchronized video with embedded audio and did not see additional actions in the OR that might interfere with the individual scene perception.

Trainee feedback and didactic results

We broadcasted three AR-based 3D videos of CI surgeries in January, September, and November 2020. The system scaled up to five different remote locations in parallel in two countries: The Netherlands (TU Delft, Rotterdam/Erasmus MC) and Germany (Fraunhofer HHI Berlin, Ludwig Maximilian University of Munich, Charité – Universitätsmedizin Berlin). The didactic component was taught by surgeons (63%), radiologists (21%), and anesthesiologists (4%) as well as scientists (12%). The workshop was attended by ENT specialists, physicians in training, other healthcare professionals, biomedical engineers, and medical students. In total, more than 150 persons were trained due to the easy scalability of the system. Thus, TeleSTAR overcomes the criticism that the number of persons enrolled in XR healthcare teaching studies is often small.[48]

All participants were able to capture data in the sense of an XR representation such as patient vital signs, preoperative imaging (e.g. CT and MRI), intraoperative distance, and functional measurements (e.g. electrocochleography),
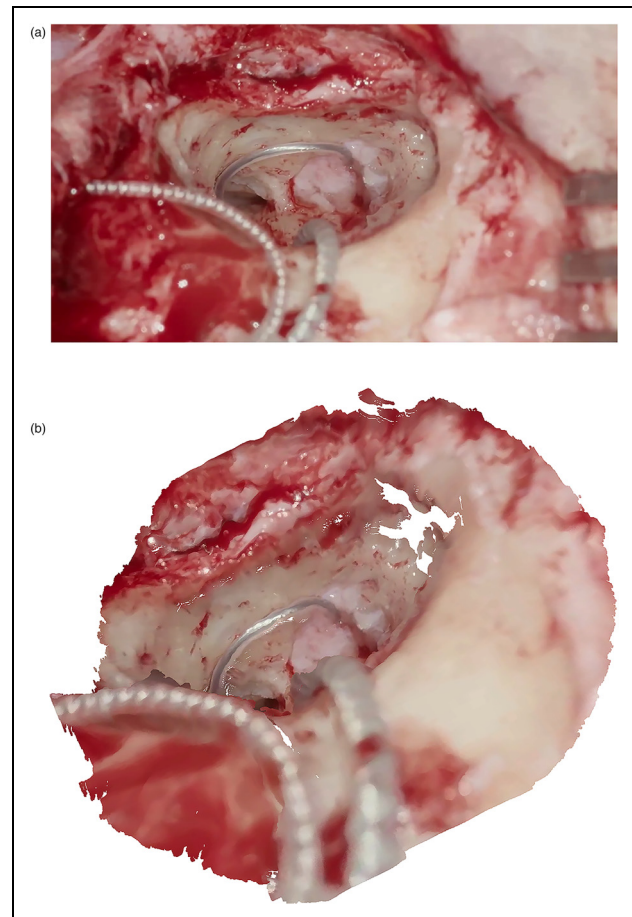
Figure 12. Comparison of 2D image and corresponding 3D reconstruction. (a) Left view of stereoscopic image pair used for 3D reconstruction. (b) Dense reconstructed point cloud of the surgical scene during a CI insertion. CI: cochlear implant; 2D: two-dimensional; 3D: three-dimensional.

or tissue analysis. This additional information helped to explain surgical steps and methods. Basic comprehension questions, such as size ratios and tissue types, could be answered easily. The possibility to annotate anatomical regions is a key feature for remote mentoring leading to more detailed questions about the procedure and its challenges. Furthermore, all participants were able to adapt fully to the surgeon's view even in surgically demanding situations and ask relevant questions to the moderating surgeon immediately. This leads to a deeper insight into the action of an experienced surgeon so that essential aspects can be taught faster and more transparent.

To evaluate the training course, questionnaires were prepared for the participants consisting of 19 questions based on Weiss et al.[49] These six questions were extended by 10 other questions to optimally address the training effect of the AR, possible future features and the technical knowledge of the participants. In total, 62 of 82 questionnaires were answered by the participants, whose mean age was (38.3 ± 4.5) years. Questionnaires were only returned from the sites where a 3D live transmission was provided. The 68 participants who took part via the 2D-YouTube-Livestream did not return any questionnaires. Due to the anonymity of these YouTube-participants, it was not possible to actively request the questionnaires.

Participants in the live 3D-sessions were from different professional groups: ENT doctors (45.5%), medical students (24.2%), engineers (13.1%), other physicians (6.1%), and radiologists (3%), see Figure 13. Engineers attended the live sessions for a better understanding of medical needs for future developments.

In the first course, which was mainly dedicated to test the technical setup with a limited number of participants, 10 doctors in specialist training for ENT took part in a seminar room with the described latency-free 3D-setup. The second course was enrolled to a bigger group with 33 participants at TU Delft, the Netherlands. The third course had 39 participants at the live transmission sides with 3D-setups in Berlin, Rotterdam, and Munich and was streamed in 2D via YouTube. Due to the COVID-19 pandemic, the number of participants was restricted at all live transmission sites.

The questionnaire (see the Appendix) was divided into questions about the training effect, possible future features, and the technical knowledge of the participants. Each question could be answered by one-of-five ratings ("strongly agree", "agree", "unsure", "disagree", "strongly disagree") (Figures 14 to 16).

The question whether stereo visualization is helpful to better perceive anatomy compared to a two-dimensional view was answered with "agree" or "strongly agree" in
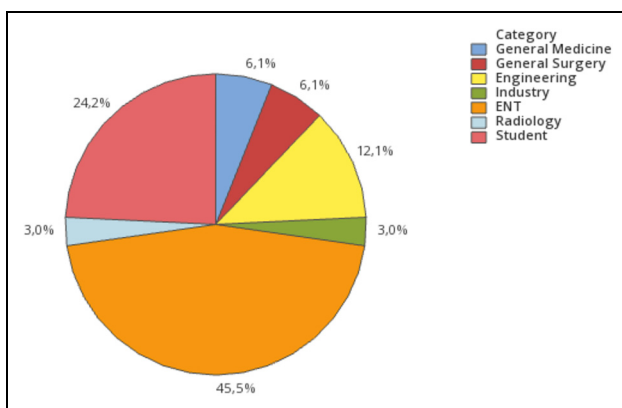

Figure 13. Professional groups of the registered participants.

58 cases, while two participants were "undecided" and two "disagreed". Similarly, in the question whether stereo visualization is helpful to understand the course of preparation, 48.3% strongly agreed, 34.5% agreed, 13.8% were undecided and 3.4% disagreed. However, 60 participants strongly agreed or agreed that stereo visualization delivers a didactic value. The annotations were found to be a valuable tool (88.7%), while 8.1% of the participants were not sure. The comprehensive questionnaire also showed that interaction with a moderator in the OR or even with other trainees in the course is very valuable (Figure 14).

In the part of the questionnaire where we assessed additional possibilities for future training courses (Figure 15) it was shown, that most participants find the possibility to attend an online course from home attractive. Only 3.2% strongly disagreed and 8.1% were undecided. Nearly the same was shown for an online-course from home in 3D with additional annotations in the video stream where 53.2% strongly agreed, 35.5% agreed, 4.8% were undecided, and 1.6% completely disagreed. However, HMDs as visualization devices are only seen as an option by 58% of the participants who attended the course on a large 3D screen with polarization glasses. Here 12.5% strongly agreed and 45.8% only agreed. 37% of the participants were undecided and 4% were sure, that an HMD is not an option. In contrast to the visualization system, most of the participants agreed that the video session should be archived for later training (79.3% agree; 13.8% disagree; 6.9% undecided).

The questionnaire part about the technical details (Figure 16) showed that additional presentations describing the background of the used technology are of interest in 89.5%. Only 3.5% found these unnecessary. The often-discussed discomfort in watching 3D was also investigated: 30.6% strongly agreed that they do not have any problems with 3D imaging, 43.5% agreed while 17.7% were not sure, 4.8% disagreed and 3.2% strongly disagreed. Audio and video quality of the 3D live streaming was found not sufficient by four respective two participants while in both questions six were not able to rate the quality.

It could shown, that AR has a strong didactic value (agreed in 78.9% and disagreed in 7%). In the survey, it was also demonstrated, that most of the participants do not have deep knowledge of how 3D live streams are generated, nor in XR technology. Knowledge of stereo-vision/3D was only "strongly agreed" by seven participants; "agreed" from 18 participants (31.6%) and "disagreed" from 13 participants (22.8%). Most participants were undecided (33.3%) while five left this question empty. Further, the knowledge in XR is just above the mean of the participants.

Conclusion

TeleSTAR provides a highly scalable solution for surgical education and training solving the problems and limitations
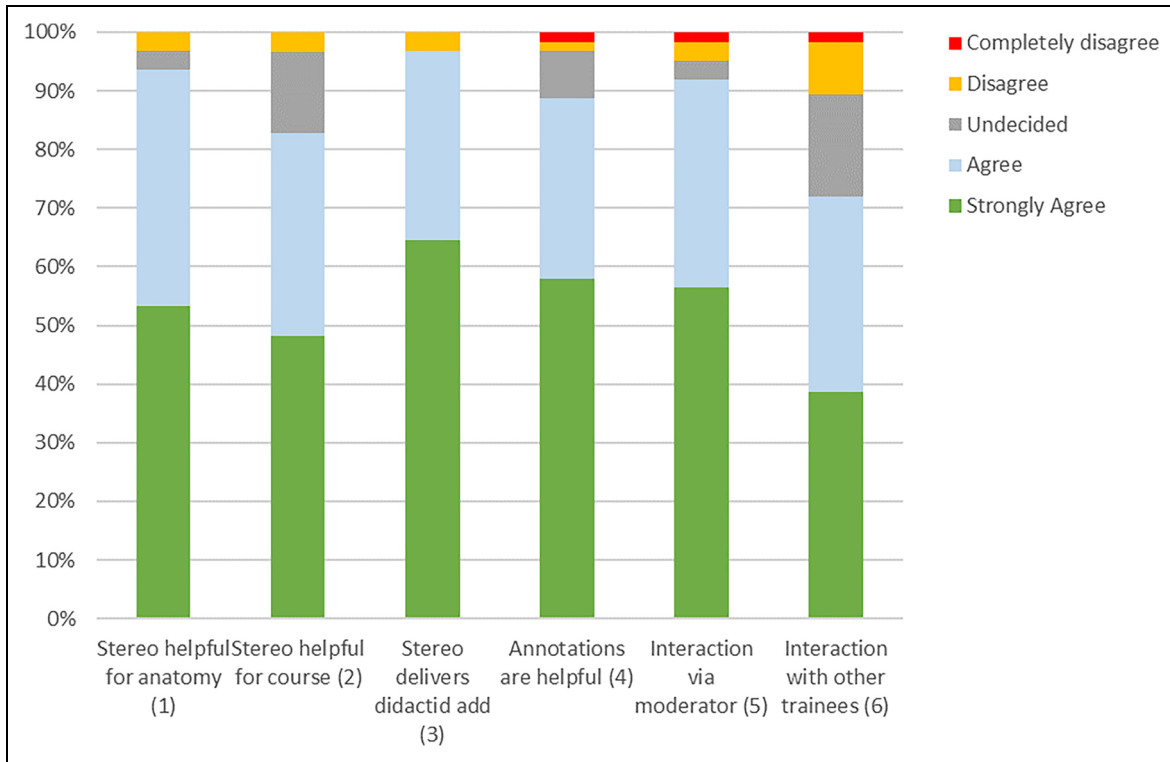
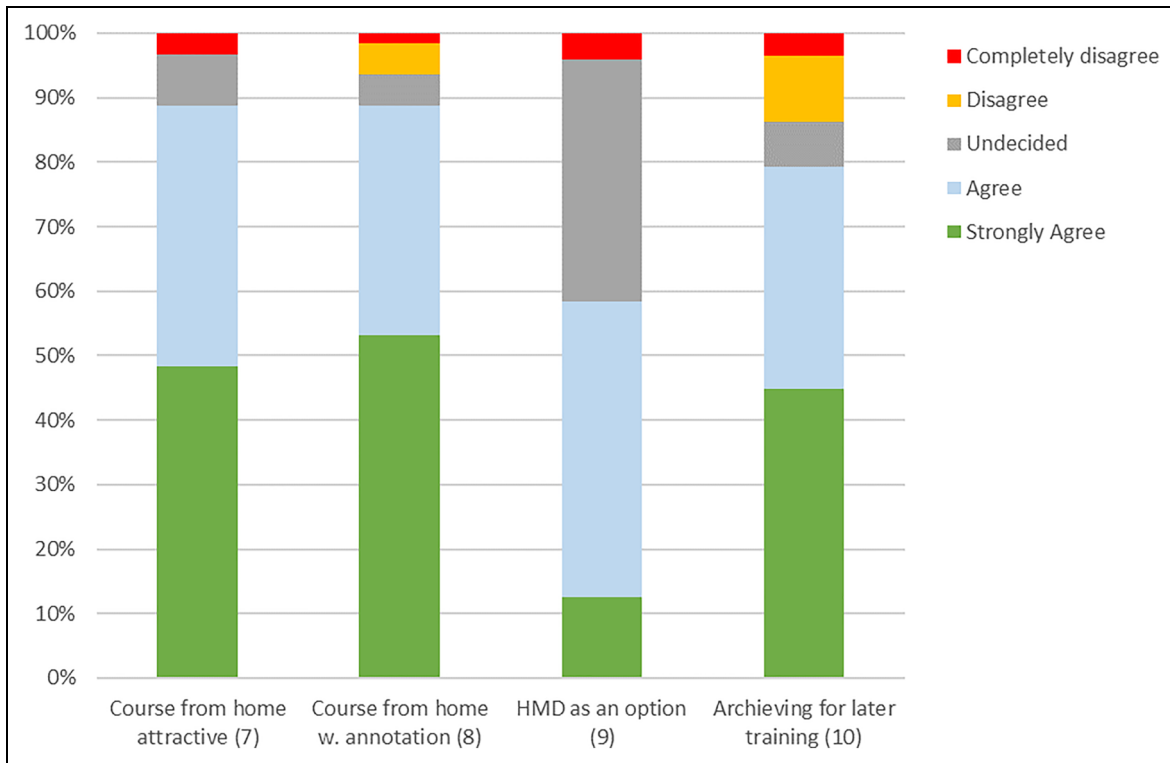**Figure 14.** Resulting answers of the questions about features used during the training session (*n* = xx).

Figure 15. Results of the questionnaire part about additional possibilities for future training courses (*n* = xx).
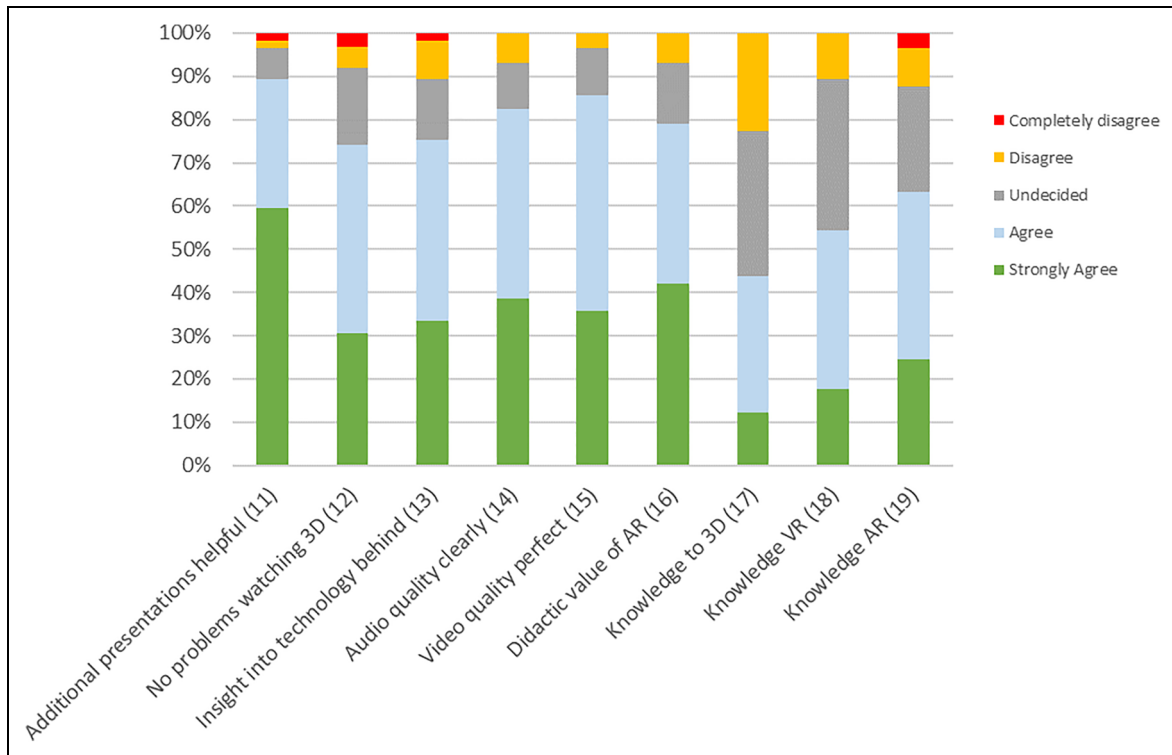
Figure 16. Results of the questionnaire part about the technology and knowledge of technical solutions.

of CME. The survey showed comparable results to other reported surgical live streaming technologies.[29] In our expanded setup, the TeleSTAR concept specifically focuses on the surgical field-of-view and improves the lack of image-based annotation and audio-visual commentary options. Trainees in different remote locations can follow a surgery as 3D live stream on displays or HMDs and are provided with important additional information using the XR-tools leading to an increased surgical transparency and direct interaction with the surgeons. This is achieved by an adaptive combination of modular software and hardware modules which guarantees a seamless way of audio-visual communication between experts and trainees. The system is scalable and allows an easy transfer to other surgical domains. In addition, TeleSTAR can also strengthen the international collaboration in surgical education.

The interactive course design promotes the direct knowledge transfer between inexperienced and experienced participants. New surgical ideas and concepts for intraoperative assistance can develop much faster with large group discussions on surgical workflows. The results highlight that our XR-based setup is a valuable tool for the current COVID-19 pandemic, but also shows great potential for surgical education in a daily routine since it has a positive impact on the learning curve of trainees.

In the future, it is contemplated to build a larger training platform to combine different aspects with scalable, adaptive, and interactive online as well as offline courses with integrated 3D/AR streaming. The key feature of simulation-based medical education is the direct feedback to the trainee based on his performance during a learning experience.[48] The modularity of our system allows an easy integration and assignment of different training tasks. Feasible tasks could be, for example, parallel estimation of tumor size during the procedure with knowledge of slice imaging or identification of pre-defined surgical landmarks. Virtual answers of trainees could be captured and evaluated to the results of the performing surgeon.[33]

Finally, the whole concept offers a new approach for clinical decision making and remote surgical education by its easy integration into an interactive processing chain. It supports intraoperative assistance from remote or on-site experts allowing discussion of complicated procedures while guaranteeing the same surgical view and consistent surgical data.

In conclusion, the proposed TeleSTAR platform presents a training platform with a high potential and provides an efficient tool for visualizing intraoperative results from medical examinations and clinical notes as well as for sharing relevant information between remote experts.

Acknowledgements

the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by Ethics Committee of Charité – Universitätsmedizin Berlin, Germany.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iDs

Eric L Wisotzky https://orcid.org/0000-0001-5731-7058
Philipp Arens https://orcid.org/0000-0001-7579-1300
Peter Eisert https://orcid.org/0000-0001-8378-4805
Armin Schneider https://orcid.org/0000-0002-0981-0478

Notes

i. ARRISCOPE, Munich Surgical Imaging, Germany.
ii. ATEM Mini, Blackmagic Design.
iii. ENC-400/DEC-400, Teracue, Germany.

References

1. De Ponti R, Marazzato J, Maresca AM et al. Pre-graduation medical training including virtual reality during COVID-19 pandemic: a report on students' perception. *BMC Med Educ* 2020; 20: 1–7.
2. Moentmann MR, Miller RJ, Chung MT et al. Using telemedicine to facilitate social distancing in otolaryngology: a systematic review. *J Telemed Telecare* 2021; 1357633X20985391.
3. Thomas EE, Haydon HM, Mehrotra A et al. Building on the momentum: sustaining telehealth beyond COVID-19. *J Telemed Telecare* 2020; 1357633X20960638.
4. Ramaswamy A, Yu M, Drangsholt S et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. *J Med Internet Res* 2020; 22: e20786.
5. Mannheim Institute of Public Health, Social and Preventive Medicine. Ärzte im zukunftsmarkt gesundheit 2020/1. Technical report, University Heidelberg, 2020.
6. Kerr B and O'leary J. The training of the surgeon: Dr. Halsted's Greatest Legacy. *Small* 1999; 96: 62.
7. Meier AH, Rawn CL and Krummel TM. Virtual reality: surgical application—challenge for the new millennium. *J Am Coll Surg* 2001; 192: 372–384.
8. Hutter MM, Kellogg KC, Ferguson CM et al. The impact of the 80-hour resident workweek on surgical residents and attending surgeons. *Ann Surg* 2006; 243: 864.
9. Scott DJ. Patient safety, competency, and the future of surgical simulation. *Simul Healthc* 2006; 1: 164–170.
10. Scott DJ, Cendan JC, Pugh CM et al. The changing face of surgical education: simulation as the new paradigm. *J Surg Res* 2008; 147: 189–193.
11. Stefanidis D, Sevdalis N, Paige J et al. Simulation in surgery: what's needed next? *Ann Surg* 2015; 261: 846–853.
12. Franklin G, Martin C, Ruszaj M et al. How the COVID-19 pandemic impacted medical education during the last year of medical school: a class survey. *Life* 2021; 11: 294.
13. Mutter D, Rubino F, Temporal MSG et al. Surgical education and internet-based simulation: the world virtual university. *Minim Invasive Ther Allied Technol* 2005; 14: 267–274.
14. De Visser H, Watson MO, Salvado O et al. Progress in virtual reality simulators for surgical training and certification. *Med J Aust* 2011; 194: S38–S40.
15. Gurusamy K, Aggarwal R, Palanivelu L et al. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *J Brit Surg* 2008; 95: 1088–1097.
16. Lahanas V, Georgiou E and Loukas C. Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators. *Int J Adv Robot Autom* 2016; 1: 1–9.
17. Lam CK, Sundaraj K and Sulaiman MN. A systematic review of phacoemulsification cataract surgery in virtual reality simulators. *Medicina* 2013; 49: 1.
18. Ruthenbeck GS and Reynolds KJ. Virtual reality for medical training: the state-of-the-art. *J Simul* 2015; 9: 16–26.
19. Cong X and Li T. Design and development of virtual medical system interface based on VR-AR hybrid technology. *Comput Math Methods Med* 2020; 2020: 0.
20. Nair L and Patel A. Mixed reality in plastic surgery: a primer. *Plast Reconstr Surg* 2018; 142: 612e–613e.
21. Wisotzky EL, Rosenthal JC, Eisert P et al. Interactive and multimodal-based augmented reality for remote assistance using a digital surgical microscope. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, pp. 1477–1484.
22. Briganti G and Le Moine O. Artificial intelligence in medicine: today and tomorrow. *Front Med (Lausanne)* 2020; 7: 27.
23. Zuo Y, Jiang T, Dou J et al. A novel evaluation model for a mixed-reality surgical navigation system: where microsoft hololens meets the operating room. *Surg Innov* 2020; 27: 193–202.
24. Kossack B, Wisotzky E, Hilsmann A et al. Automatic region-based heart rate measurement using remote photoplethysmography. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2755–2759.
25. Kossack B, Wisotzky EL, Hilsmann A et al. Local blood flow analysis and visualization from rgb-video sequences. *Curr Dir Biomed Eng* 2019; 5: 373–375.
26. Kubben PL and Sinlae RS. Feasibility of using a low-cost head-mounted augmented reality device in the operating room. *Surg Neurol Int* 2019; 10: 26.
27. Hartwig R, Ostler D, Feußner H et al. Compass: localization in laparoscopic visceral surgery. *Curr Dir Biomed Eng* 2020; 6: 20200013.
28. Mathur AS. Low cost virtual reality for medical training. In *2015 IEEE Virtual Reality (VR)*. IEEE, pp. 345–346.

29. Abu-Rmaileh M, Osborn T, Gonzalez SR et al. The use of live streaming technologies in surgery: a review of the literature. *Ann Plast Surg* 2022; 88: 122–127.

30. Mark TL, Treiman K, Padwa H et al. Addiction treatment and telehealth: review of efficacy and provider insights during the COVID-19 pandemic. *Psychiatr Serv* 2022; 73: 484–491.

31. Fritz T, Stachel N and Braun BJ. Evidence in surgical training – a review. *Innov Surg Sci* 2019; 4: 7–13.

32. Gonzalez JN, Axiotakis Jr LG, Yu VX et al. Practice of telehealth in otolaryngology: a scoping review in the era of COVID-19. *Otolaryngol Head Neck Surg* 2022; 166: 417–424.

33. Jain S, Lee S, Barber SR et al. Virtual reality based hybrid simulation for functional endoscopic sinus surgery. *IISE Trans Healthc Syst Eng* 2020; 10: 127–141.

34. Wisotzky EL, Rosenthal JC, Wege U et al. Surgical guidance for removal of cholesteatoma using a multispectral 3D-endoscope. *Sensors* 2020; 20: 5334.

35. Gard N, Rosenthal JC, Jurk S et al. Image-based measurement by instrument tip tracking for tympanoplasty using digital surgical microscopy. In Fei B and Linte CA (eds.) *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951. International Society for Optics and Photonics, SPIE, pp. 318–328. doi:10.1117/12.2512415.

36. Wisotzky EL, Uecker FC, Rosenthal JC et al. Near-UV to near-IR multispectral illumination in a digital surgical microscope. *Curr Dir Biomed Eng* 2021; 7: 464–467.

37. Liu GS, Shenson JA, Farrell JE et al. Signal to noise ratio quantifies the contribution of spectral channels to classification of human head and neck tissues ex vivo using deep learning and multispectral imaging. *J Biomed Opt* 2023; 28: 016004.

38. Leonhardt M and Brendel H. Critical spectra in the color reproduction process of digital motion picture cameras. In *Color and Imaging Conference*. 1, Society for Imaging Science and Technology, pp. 167–170.

39. Clark C, Reisner D, Stump D et al. Report from the American Society of Cinematographers Technology Committee. *SMPTE Motion Imaging J* 2013; 122: 46–53.

40. Wisotzky EL, Kossack B, Uecker FC et al. Validation of two techniques for intraoperative hyperspectral human tissue determination. *J Med Imaging* 2020; 7: 065001.

41. Allan M, Mcleod J, Wang C et al. Stereo correspondence and reconstruction of endoscopic data challenge, 2021.

42. Rosenthal J, Gard N, Schneider A et al. Kalibrierung stereoskopischer systeme für medizinische messaufgaben. In *Proceedings of the 16th Annual Conference of the German Society for Computer and Robotic Assisted Surgery (CURAC)*, 2017. pp. 161–163.

43. Hope C, Reilly JJ, Griffiths G et al. The impact of COVID-19 on surgical training: a systematic review. *Tech Coloproctol* 2021; 25: 505–520.

44. Marketresearchcom. Cochlear implant market size, share & trends analysis report by type of fitting (unilateral implants, bilateral implants), by end-use (adult, pediatric), by region, and segment forecasts, 2021 - 2028. Technical Report GVR-1-68038-720-9, Grand View Research, 2021.

45. Roosli C, Sim JH, Möckel H et al. An artificial temporal bone as a training tool for cochlear implantation. *Otol Neurotol* 2013; 34: 1048–1051.

46. Fischer HG, Zehlicke T, Gey A et al. Multimodales weiterbildungskonzept schläfenbeinchirurgie. *HNO* 2021; 69: 545–555.

47. Schneider A, Lanski M, Bauer M et al. An AR-solution for education and consultation during microscopic surgery. *Int J Comput Assist Radiol Surg* 2019; 14: S59–S60.

48. Viglialoro RM, Condino S, Turini G et al. Augmented reality, mixed reality, and hybrid approach in healthcare simulation: a systematic review. *Appl Sci* 2021; 11: 2338.

49. Weiss NM, Schneider A, Hempel JM et al. Evaluating the didactic value of 3D visualization in otosurgery. *Eur Arch Otorhinolaryngol* 2021; 278: 1027–1033.

Appendix

1. Timestamp 1: 1–10 min
   (a) After disinfection infiltration of xylonest with adrenalin
   (b) Retroauricular skin incision
   (c) Preparation of the muscle-periosteal flap and extraction of fascia for later transplantation
   (d) **AR-Feature:** 1st usage 3D reconstruction mode in overview mode
2. Timestamp 2: 10–35 min
   (a) Exposure of the planum mastoideum and subtle control of bleedings (hemostasis)
   (b) Mastoidectomy with presentation of the short incus appendix and semicircular canal
   (c) Creating and displaying of Wullstein's window and the two nerves with the chorda-facialis angle until the complete opening until the stapedial tendon and the middle ear structures are clearly visible
   (d) **AR-Feature:** 2nd usage 3D reconstruction mode
   (e) **AR-Feature:** 1st usage annotation mode
   (f) **AR-Feature:** 1st usage spectral mode
   (g) Representation of the round window
3. Timestamp 3: 35–40 min
   (a) Drilling of the implant site in the bony skull to place the CI
4. Timestamp 4: 40–70 min
   (a) Complete electrode insertion
   (b) Subtle sealing of the electrode entry point with pieces of fascia
   (c) Regulation-compliant CI device testing (impedance measurements) showing as picture-in-picture
   (d) Verification of acoustic/stapedius reflex
   (e) **AR-Feature:** 3rd usage 3D reconstruction mode
   (f) **AR-Feature:** 2nd usage annotation mode
   (g) Telemetric derivation of the potentials
   (h) Re-implantation of fasciae and possibly muscle grafts to fixate the electrode cable

5. Timestamp 5: 70–80 min
   (a) Re-implantation of intraoperatively collected bone meal for placement on the electrode array and the bony canal running to the implant
   (b) Surgical wound closure
   (c) End of Procedure

Feedback Questionnaire To improve surgical training for the future and to improve further upcoming TeleSTAR courses, we would like to ask you to answer some questions. Do you agree with the following statements?

1. Thanks to the 3D/stereo representation of the surgical field, I can perceive the anatomical topography and structures better than with a conventional 2D representation.

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

2. I can follow the course of the preparation better with the 3D/stereo representation of the surgical field than with a 2D representation.

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

3. The possibility to see the surgical field as a co-observer in 3D/stereo provides a didactic added value for surgical courses.

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

4. The possibility of integrating graphical annotations in the video of the surgical field add didactic value for surgical courses.

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

5. The possibility to see the operation live via 3D/stereo video transmission from home would make an online distance-learning course very attractive.

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

6. The interaction with the surgeon via a moderator and the annotation mode (bi-directional?) would provide additional didactic value!

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

7. The interaction with other trainees using a lecture room-based annotation mode would provide additional didactic value!

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

8. The possibility to see the operation with annotation live via 3D/stereo video transmission from home would make an online distance-learning course even more attractive!

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

9. What do you think of Head-Mount-Displays (AR/VR-glasses) to watch a surgery in 3D?

   ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

10. The 3D/stereo video data of a surgery should be archived for self-study and made available online for registered users.

    ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

11. Additional medical lectures/presentations to the live surgery are helpful for a better understanding what and how the surgery is performed?

    ☐ Strongly agree ☐ Agree ☐ Unsure ☐ Disagree ☐ Strongly disagree

12. I have no problems watching 3D/stereo movies and videos (e.g. discomfort, dizziness, headaches).

☐ ☐ ☐ ☐ ☐
Strongly agree   Agree   Unsure   Disagree   Strongly disagree

13. It is interesting to get an insight in the technology how image capturing and video transmission with the digital surgical microscope is realized.

☐ ☐ ☐ ☐ ☐
Strongly agree   Agree   Unsure   Disagree   Strongly disagree

14. The audio quality of the transmission allowed me to understand everything clearly.

☐ ☐ ☐ ☐ ☐
Strongly agree   Agree   Unsure   Disagree   Strongly disagree

15. Video quality of the transmission was clear and without disturbing artefacts.

☐ ☐ ☐ ☐ ☐
Strongly agree   Agree   Unsure   Disagree   Strongly disagree

16. How do you rate the didactic value of the presented AR features—Annotation mode and depth visualisation?

☐ ☐ ☐ ☐ ☐
Very helpful   helpful   undecided   Not helpful   distracting

17. What previous knowledge do you have regarding 3D/VR/AR—(1: no knowledge–5: Expert)?

	1	2	3	4	5
3D	☐	☐	☐	☐	☐
AR	☐	☐	☐	☐	☐
VR	☐	☐	☐	☐	☐

18. Which 3D/AR visualization features could be also useful for surgical training?

_____
_____

19. Do you have any more comments/suggestions?

_____
_____